Tools for Recognizing Useful Signals of Trustworthiness (TRUST)

The INSTINCT Challenge: Predicting Trustworthiness From Others' Signals

Program Manager: Dr. Adam Russell; E-mail: adam.russell@iarpa.gov

I.N.S.T.I.N.C.T.

Overview

The goal: predict one person's trustworthiness from another's signals

The Method: Statisticians, programmers compete to get the most from neural, physiological, & behavioral data

The Results: 15% improvement over baseline analysis

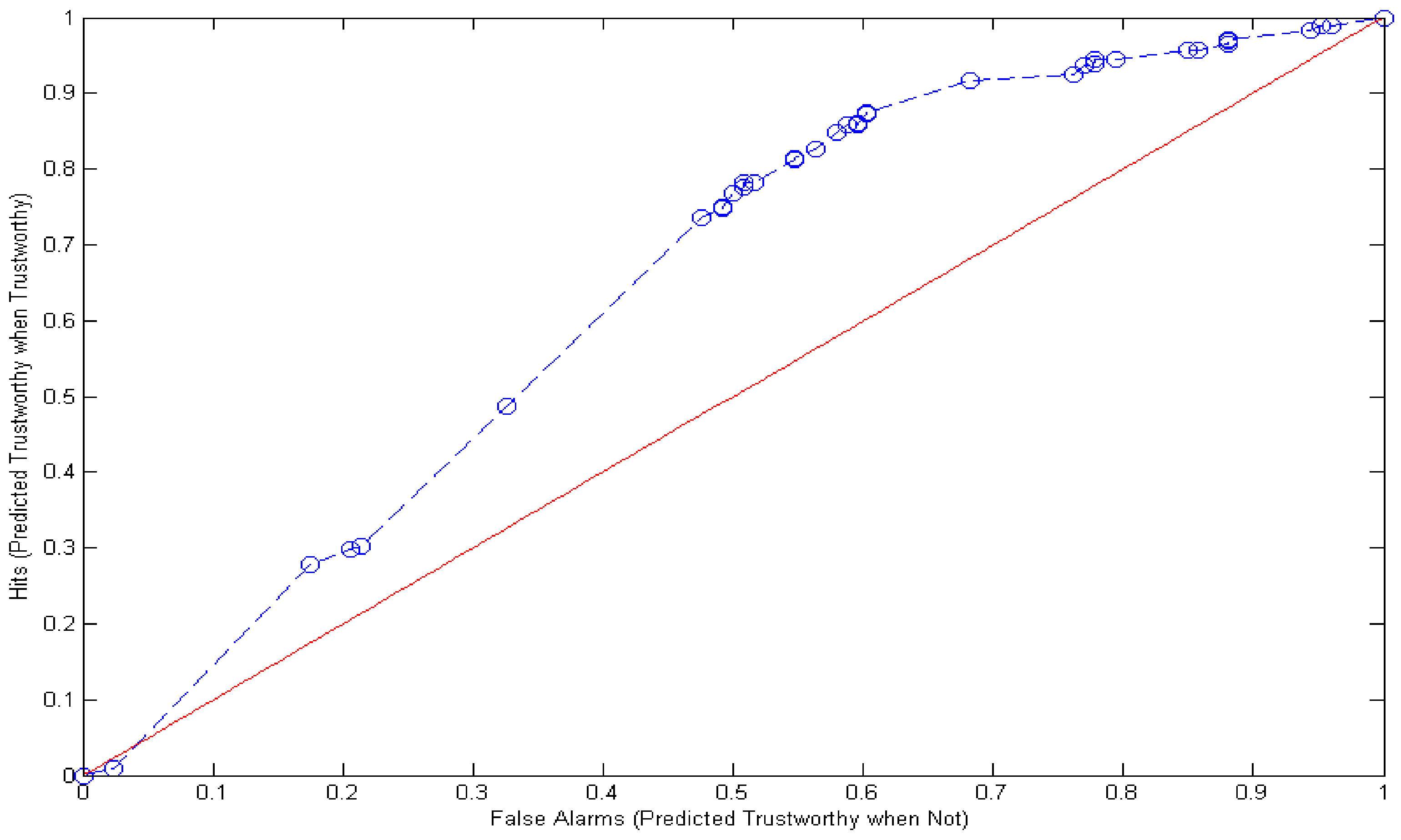
Key Findings: Heart rate and response time are more predictive than EEG and hormone data

Winning Solution

JEDI MIND: Joint Estimation of Deception Intent via Multisource Integration of Neurophysiological Discriminators

Performance: 15% improvement over baseline (Maximum d' = .9, reflecting 92% hits & 68% false alarms)

Best Predictive Signals: heart rate, Decision response time



INSTINCT Challenge Details

Dates: Open February 19-May 5, 2014

Registered Solvers: 453

Submitted Algorithms: 39; 7 finalists evaluated against additional data

Data Set: TRUST Program experiments in which volunteers made promises to each other & chose whether to keep them, for monetary stakes

Requirements: Beat baseline analysis with d' > .7 (high hits, low false alarms).

Conclusions and Next Steps

- Self-produced signals *can* improve prediction of others
- Significant advances are needed to lower false alarms for practical use.
- Current neural and hormonal metrics underperformed—but behavioral data still have untapped predictive potential
- New analytic methods are powerful, but subject matter expertise still matters—finalists depending on stats alone didn't hold up
- Next steps: test JEDI MIND on other data sets; potential RFI on self-as-sensor research

Why a Challenge?

- Try multiple approaches and expertises in parallel
- Low-risk testing of high-risk efforts
- Raise awareness of problem area